DANNY SIU

https://dannysiudata.com
(817) 320-7283
dsiucareer@gmail.com

EXPERIENCE

Smarsh *Remote*
Software Engineer Jun. 2021 - Jul. 2022

- Released new features for Smarsh's Conduct Surveillance application, which employs pre-trained NLP and ML models to identify misconduct in regulated electronic and voice communications and monitors 8B+ messages from many major global financial institutions.
- Resolved software bugs and production problems escalated from the support team, which evoked company recognition of my team and improved relationships with at-risk clients.

University of California at Berkeley, College of Chemistry *Berkeley, CA*
Research Intern for the Cohen Research Group Feb. 2019 - Aug. 2019

- Launched efforts to replace a computationally expensive WRF-CHEM climate model run with ML model predictions.
- Began preparations to produce higher resolution measurements of NO2 levels than measured by NASA's Aura satellite.
- Aligned 47K WRF-CHEM pixels with Aura's OMI pixels to enable better data correlation.

University of Texas at Austin *Austin, TX*
Summer Student Researcher Jun. 2016 - Aug. 2016

- Created molecular visualizations of catalyst binding to communicate research on energy production efficiency.
- Computationally simulated carbon monoxide binding on alloys of gold-palladium nanoparticles during ORR-reactions.

PROJECTS

More on my website: **https://dannysiudata.com**

POS-Tagger Model Python (PyTorch, NumPy)

- Implemented and trained an LSTM for POS-tagging by using mini-batch stochastic gradient descent on pre-trained GloVe embeddings.
- Improved model accuracy by implementing a bi-directional LSTM and dropout layer, and loading the 500,000 most common 300-dimensional word embeddings.

Movie Review Sentiment Classifier Python (Scikit-Learn, SciPy, NumPy)

- Created a binary sentiment classifier for movie reviews using featurization of bag-of-words, bigrams, trigrams, sentiment dictionaries, and neutral word removal.

Spam Email Classifier Python (Scikit-Learn, Matplotlib, Pandas, Seaborn, NumPy)

- Performed exploratory data analysis on textual data to perform feature selection.
- Implemented feature engineering for a logistic regression model and utilized carefully selected spam keywords to fit my spam classifier model to achieve 95% accuracy on the validation set and test set.

SKILLS

Languages: Python, Java, SQL, PostgreSQL, C, HTML, CSS, PyTorch, R
Tools and Frameworks: Linux, Git, Jupyter notebooks, Agile, Kanban, LaTeX, Bash, Relational DBMS
Data modeling experience: Classification, Prediction, Sentiment analysis, Clustering, Data visualization, Neural networks, Feature engineering, Data analysis, Data mining, Machine learning algorithms

EDUCATION

University of California at Berkeley *Berkeley, CA*
B.A. Data Science; Minor in Computer Science; Domain Emphasis in Linguistics; Aug. 2017 - May 2021
National Merit Scholar; Valedictorian of Lamar High School

Relevant Coursework: Algorithms, Natural Language Processing, Machine Learning & Data Analytics, Data Structures, Data Mining and Analytics, Discrete Math & Probability Theory, Computer Architecture, Data Engineering, Linux System Administration, Linear Algebra & Differential Equations